

USING INTRA-DOCUMENT INDICES TO IMPROVE XQUERY PROCESSING OVER XML STREAMS

ABSTRACT

5

A system and method for parsing documents in query processing comprises producing at least one index of a document written in a mark-up language, corresponding the index to the document, scanning the document, and selectively skipping portions of the document based on instructions from the index. Furthermore, the mark-up language comprises any of HTML and XML; the skipped portions of the document comprise portions irrelevant to the query; the index comprises a plurality of elements representing textual categories of the query; and the instructions match the elements to the query. If the elements do not match the query, then the parser uses the index to skip the portions of the document corresponding to the unmatched elements. Moreover, each of the elements corresponds to a position in the document, wherein the position comprises an end position, which determines where to resume scanning the document upon skipping the portions of the document.

10

15